Comput Optim Appl (2013) 56:317–341
DOI 10.1007/s10589-013-9558-3

# Symmetric Perry conjugate gradient method

**Dongyi Liu · Genqi Xu**

Received: 11 May 2011 / Published online: 29 March 2013


**Abstract** A family of new conjugate gradient methods is proposed based on Perry's idea, which satisfies the descent property or the sufficient descent property for any line search. In addition, based on the scaling technology and the restarting strategy, a family of scaling symmetric Perry conjugate gradient methods with restarting procedures is presented. The memoryless BFGS method and the SCALCG method are the special forms of the two families of new methods, respectively. Moreover, several concrete new algorithms are suggested. Under Wolfe line searches, the global convergence of the two families of the new methods is proven by the spectral analysis for uniformly convex functions and nonconvex functions. The preliminary numerical comparisons with CG_DESCENT and SCALCG algorithms show that these new algorithms are very effective algorithms for the large-scale unconstrained optimization problems. Finally, a remark for further research is suggested.

**Keywords** Conjugate gradient method · Descent property · Spectral analysis · Global convergence

This research is supported by the Natural Science Foundation of China grant NSFC-61174080 and by the Seed Foundation of Tianjin University.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10589-013-9558-3) contains supplementary material, which is available to authorized users.

D. Liu (✉) · G. Xu
Department of Mathematics, Tianjin University, Tianjin 300072, P.R. China
e-mail: dyliu@tju.edu.cn

G. Xu
e-mail: gqxu@tju.edu.cn

## 1 Introduction

The classical conjugate gradient (CG) method with line search is as follows:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1}$$

where the directions $d_k$ is given by

$$\begin{cases} d_1 = -g_1, \\ d_{k+1} = -g_{k+1} + \beta_k d_k, \quad \forall k \geq 1, \end{cases} \tag{2}$$

where $g_k = g(x_k) = \nabla f(x_k)$. The different choices for the parameter $\beta_k$ correspond to different CG methods, such as HS method [15], FR method [7], PRP method [22, 23], LS method [16], PRP$^+$ method [8], DY method [5] and so on. On the history of the conjugate gradient method, there are several survey articles, such as [11].

In [17], the Perry conjugate gradient algorithm [21] was generalized and the line search directions were formulated as follows:

$$\begin{cases} d_1 = -g_1, \\ d_{k+1} = -P_{k+1} g_{k+1} = -g_{k+1} + \frac{y_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} s_k - \sigma \frac{s_k^{\mathrm{T}} g_{k+1}}{y_k^{\mathrm{T}} u_k} u_k, \quad \forall k > 1, \end{cases} \tag{3}$$

where $s_k = x_{k+1} - x_k = \alpha_k d_k$, $y_k = g_{k+1} - g_k$, $\alpha_k$ is the steplength of the line search and $\sigma$ is a preset parameter,

$$P_{k+1} = \left( I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \sigma \frac{u_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} u_k} \right), \tag{4}$$

which is called Perry iteration matrix, and the vector $u_k$ is any vector in $\mathbb{R}^n$ such that $y_k^T u_k \neq 0$. In the paper [17], the case $u_k = y_k$ was discussed. When $u_k = s_k$, the CG_DESCENT algorithm [10–12] can be deduced and the D-L method [4] can be derived from the restriction $\sigma > 0$. Recently, we also studied the case $u_k = s_k$ in [19] and presented a RSPDCGs algorithm.

In this paper, a family of symmetric Perry conjugate gradient methods is proposed, that is, the line search directions are formulated by

$$\begin{cases} d_1 = -g_1, \\ d_{k+1} = -Q_{k+1} g_{k+1} = -g_{k+1} + \beta_k d_k + \gamma_k y_k, & \forall k \geq 1 \\ \qquad = -g_{k+1} + \frac{s_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} y_k + \left[ \frac{y_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} - \left( \sigma + \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k} \right) \frac{s_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} \right] s_k, & \forall k \geq 1 \end{cases} \tag{5}$$

where $\beta_k = \frac{y_k^{\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k} - (\alpha_k \sigma + \frac{y_k^{\mathrm{T}} y_k}{d_k^{\mathrm{T}} y_k}) \frac{d_k^{\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k}$, $\gamma_k = \frac{d_k^{\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k}$ and

$$Q_{k+1} = I - \frac{s_k y_k^{\mathrm{T}} + y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \left( \sigma + \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k} \right) \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}, \tag{6}$$

which is called the symmetric Perry iteration matrix. When $\sigma y_k^{\mathrm{T}} s_k > 0$, for any line search, the directions defined by (5) satisfy the descent property [1]

$$d_{k+1}^{\mathrm{T}} g_{k+1} < 0, \tag{7}$$

or the sufficient descent property [8]

$$d_{k+1}^{\mathrm{T}} g_{k+1} \leq -c_0 \|g_{k+1}\|^2 \quad (c_0 > 0). \tag{8}$$

This paper is organized as follows. In Sect. 2, first, the family of the symmetric Perry conjugate gradient methods is deduced. Then the spectra of the iteration matrix are analyzed, so, its sufficient descent property is proved and several concrete algorithms are proposed. In Sect. 3, the scaling technology and the restarting strategy are applied to the symmetric Perry conjugate gradient methods, thus, a family of scaling Perry conjugate gradient methods with restarting procedures is developed. In Sect. 4, the global convergence of the two families of the new methods with the Wolfe line searches is proven by the spectral analysis of the conjugate gradient iteration matrix. In Sect. 5, the preliminary numerical results are reported. A remark for further research is given in Sect. 6.

## 2 The symmetric Perry conjugate gradient method

In [21], A. Perry changed the CG update parameter $\beta_k$ of the HS conjugate gradient method [15] into $\beta_k^P = \frac{(y_k - s_k)^{\mathrm{T}} g_{k+1}}{y_k^{\mathrm{T}} d_k}$, and formulated the line search directions

$$d_{k+1} = -g_{k+1} + \beta_k^P d_k = -Q_{k+1} g_{k+1}, \quad k = 1, 2, \ldots,$$

and

$$y_k^{\mathrm{T}} d_{k+1} = -s_k^{\mathrm{T}} g_{k+1}, \tag{9}$$

where

$$Q_{k+1} = D_{k+1} + \frac{s_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k} \quad \text{and} \quad D_{k+1} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right). \tag{10}$$

In [17], (10) and (9) were substituted by

$$Q_{k+1} = D_{k+1} + u v^{\mathrm{T}} \tag{11}$$

and

$$y_k^{\mathrm{T}} d_{k+1} = -\sigma s_k^{\mathrm{T}} g_{k+1}, \tag{12}$$

respectively, where $u, v \in \mathbb{R}^n$ and $\sigma$ is a parameter. Thus, it is follows from (11), (12) and $d_{k+1} = -Q_{k+1} g_{k+1}$ that $(\sigma s_k - v y_k^{\mathrm{T}} u)^{\mathrm{T}} g_{k+1} = 0$, which yields $v = \frac{\sigma s_k}{y_k^{\mathrm{T}} u}$. So,

$$Q_{k+1} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \sigma \frac{u s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} u}\right), \tag{13}$$

from which the generalized Perry conjugate gradient method ((1) and (3)) can be obtained [17].

In this paper, we choose a suitable $u$ such that $Q_{k+1}$ is a symmetric matrix, thus the line search directions $d_k$ may satisfy (7) or (8). Let $Q_{k+1} = Q_{k+1}^{\mathrm{T}}$, then

$$\left(\frac{y_k}{s_k^{\mathrm{T}} y_k} + \sigma \frac{u}{y_k^{\mathrm{T}} u}\right) s_k^{\mathrm{T}} = s_k \left(\frac{y_k}{s_k^{\mathrm{T}} y_k} + \sigma \frac{u}{y_k^{\mathrm{T}} u}\right)^{\mathrm{T}}.$$

Therefore, the vector $u$ can be taken as

$$u = a\left(\frac{\sigma s_k^{\mathrm{T}} y_k + y_k^{\mathrm{T}} y_k}{(s_k^{\mathrm{T}} y_k)^2} s_k - \frac{1}{s_k^{\mathrm{T}} y_k} y_k\right) \quad (a \neq 0), \tag{14}$$

and $u^{\mathrm{T}} y_k = a\sigma$. We note that the matrix $Q_{k+1}$ defined by (13) is independent of the nonzero constant $a$, so, we can choose $a = 1$. Thus, from (13) and (14) we can obtain the matrix $Q_{k+1}$ defined by (6).

The method formulated by (1) and (5) is called the symmetric Perry conjugate gradient method, denoted by SPCG. And the directions generated by (5) are called the symmetric Perry conjugate gradient directions, which will be proven to be descent directions in Sect. 2.2.

From the above discussions, a family of new nonlinear conjugate gradient algorithms can be obtained as follows:

**Algorithm 1** (SPCG)

Step 1. Give an initial point $x_1$ and $\varepsilon \geq 0$. Set $k = 1$.
Step 2. Calculate $g_1 = g(x_1)$. If $\|g_1\| \leq \varepsilon$ then stop, otherwise let $d_1 = -g_1$.
Step 3. Calculate steplength $\alpha_k$ with line searches.
Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$.
Step 5. Calculate $g_{k+1} = g(x_{k+1})$. If $\|g_{k+1}\| \leq \varepsilon$ then stop.
Step 6. Calculate the directions $d_{k+1}$ via (5) with different $\sigma$.
Step 7. Set $k = k + 1$, then go to step 3.

*Remark 1* In this paper, to ensure the convergence of the algorithm, we adopt the Wolfe line search strategies:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + b_1 \alpha_k d_k^{\mathrm{T}} g_k \tag{15}$$

and

$$d_k^{\mathrm{T}} g(x_k + \alpha_k d_k) \geq b_2 d_k^{\mathrm{T}} g_k, \tag{16}$$

where $0 < b_1 < b_2 < 1$. The stopping criterion, $\|g_k\| \leq \varepsilon$, can be changed into other forms. For the different choices of $\sigma$, several concrete forms of the algorithm will be discussed in the Sect. 2.2.

## 2.1 Spectral analysis

Here, we analyze the spectra of the Perry matrix and the symmetric Perry matrix.

**Theorem 1** *Let $P_{k+1}$ be defined by (4). Then when $\sigma(y_k^{\mathrm{T}}s_k)\neq 0$, $P_{k+1}$ is a nonsingular matrix and the eigenvalues of $P_{k+1}$ consist of* 1 *(*$n-2$ *multiplicity*), $\lambda_1^{k+1}$ *and* $\lambda_2^{k+1}$, *where*

$$\lambda_1^{k+1}=\frac{1}{2}\left[\left(1+\sigma\frac{s_k^{\mathrm{T}}u_k}{y_k^{\mathrm{T}}u_k}\right)-\sqrt{\left(1+\sigma\frac{s_k^{\mathrm{T}}u_k}{y_k^{\mathrm{T}}u_k}\right)^2-4\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}}\right] \tag{17}$$

*and*

$$\lambda_2^{k+1}=\frac{1}{2}\left[\left(1+\sigma\frac{s_k^{\mathrm{T}}u_k}{y_k^{\mathrm{T}}u_k}\right)+\sqrt{\left(1+\sigma\frac{s_k^{\mathrm{T}}u_k}{y_k^{\mathrm{T}}u_k}\right)^2-4\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}}\right]. \tag{18}$$

*Proof* From the fundamental algebra formula

$$\det\big(I+xy^{\mathrm{T}}+uv^{\mathrm{T}}\big)=\big(1+y^{\mathrm{T}}x\big)\big(1+v^{\mathrm{T}}u\big)-\big(x^{\mathrm{T}}v\big)\big(y^{\mathrm{T}}u\big),$$

it follows that

$$\det(P_{k+1})=\det\left(I-\frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}}y_k}+\frac{\sigma u_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}}u_k}\right)=\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}. \tag{19}$$

Therefore, the Perry matrix (4) is a nonsingular matrix when $\sigma y_k^{\mathrm{T}}s_k\neq 0$.

Since $\forall\xi\in\operatorname{span}\{s_k,y_k\}^{\perp}\subset\mathbb{R}^n$,

$$P_{k+1}\xi=\left(I-\frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}}y_k}+\sigma\frac{u_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}}u_k}\right)\xi=\xi-\frac{y_k^{\mathrm{T}}\xi}{s_k^{\mathrm{T}}y_k}s_k+\sigma\frac{s_k^{\mathrm{T}}\xi}{y_k^{\mathrm{T}}u_k}u_k=\xi,$$

the matrix $P_{k+1}$ has the eigenvalue 1 ($n-2$ multiplicity), corresponding to the eigenvectors $\xi\in\operatorname{span}\{s_k,y_k\}^{\perp}$.

By the relationships between the trace and the eigenvalues of matrix and between the determinant and the eigenvalues of matrix, the other two eigenvalues are the roots of the following quadratic polynomial

$$\lambda^2-\left(1+\sigma\frac{s_k^{\mathrm{T}}u_k}{y_k^{\mathrm{T}}u_k}\right)\lambda+\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}=0. \tag{20}$$

Thus, the other two eigenvalues are determined by (17) and (18), respectively. □

According to Theorem 1, the following theorem for the symmetric Perry matrix $Q_{k+1}$ defined by (6) can be deduced.

**Theorem 2** *Let $\lambda_{\min}^{(k+1)}$ and $\lambda_{\max}^{(k+1)}$ be the minimum and maximum eigenvalues of $Q_{k+1}$, respectively, where $Q_{k+1}$ is defined by (6). If $\sigma(y_k^{\mathrm{T}}s_k)>0$, then*

$$\lambda_{\min}^{(k+1)}=\frac{1}{2}\left[\omega_k+\sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}-\sqrt{\left(\omega_k+\sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right)^2-4\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}}\right], \tag{21}$$

$$\lambda_{\max}^{(k+1)} = \frac{1}{2}\left[\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} + \sqrt{\left(\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right)^2 - 4\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}}\right], \tag{22}$$

$$\frac{\sigma y_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}y_k + \sigma y_k^{\mathrm{T}}s_k} \le \lambda_{\min}^{(k+1)} \le 1 \le \omega_k \le \lambda_{\max}^{(k+1)} \le \omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} \tag{23}$$

*and*

$$\lambda_{\min}^{(k+1)} \le \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} \le \max\left\{\omega_k, \omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} - 1\right\} \le \lambda_{\max}^{(k+1)}, \tag{24}$$

*where* $\omega_k = \frac{y_k^{\mathrm{T}}y_k s_k^{\mathrm{T}}s_k}{(s_k^{\mathrm{T}}y_k)^2}$. *Moreover,* $Q_{k+1}$ *is a symmetric positive definite matrix when* $\sigma(y_k^{\mathrm{T}}s_k) > 0$.

*Proof* When $\sigma(y_k^{\mathrm{T}}s_k) > 0$, from (14), (17), (18) and the following relations:

$$\left(\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right)^2 - 4\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k} = \left(\omega_k - \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right)^2 + 4\sigma(\omega_k - 1)\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} \ge 0, \tag{25}$$

it can be proven that $\lambda_{\min}^{(k+1)}$ and $\lambda_{\max}^{(k+1)}$ are formulated by (21) and (22), respectively.
Simple calculation claims that

$$\lambda_{\max}^{(k+1)} \le \frac{1}{2}\left[\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} + \sqrt{\left(\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right)^2}\right] = \omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}$$

and

$$\begin{aligned}\lambda_{\max}^{(k+1)} &= \frac{1}{2}\left[\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} + \sqrt{4(\omega_k - 1) + \left(\omega_k - 2 + \sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}\right)^2}\right] \\ &\ge \frac{1}{2}\left[\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} + \left|\omega_k - 2 + \sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}\right|\right].\end{aligned} \tag{26}$$

So, the inequality (26) implies that

$$\lambda_{\max}^{(k+1)} \ge \omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} - 1 \ge \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}.$$

In addition, it follows from (25) that

$$\lambda_{\max}^{(k+1)} \ge \frac{1}{2}\left(\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} + \left|\omega_k - \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right|\right) \ge \omega_k.$$

Therefore, $\lambda_{\max}^{(k+1)} \ge \max\{\omega_k, \omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} - 1\} \ge \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}$.

Similarly,

$$\lambda_{\min}^{(k+1)} = \frac{1}{2}\left[\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} - \sqrt{\left(\omega_k - \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right)^2 + 4(\omega_k - 1)\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}}\right]$$
$$\leq \frac{1}{2}\left[\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} - \left|\omega_k - \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right|\right] \leq \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}$$

and

$$\lambda_{\min}^{(k+1)} = \frac{1}{2}\left[\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} - \sqrt{4(\omega_k - 1) + \left(\omega_k - 2 + \sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}\right)^2}\right]$$
$$\leq \frac{1}{2}\left[\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} - \left|\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k} - 2\right|\right] \leq 1.$$

In the end, it follows from (20) that $\lambda_{\min}^{(k+1)}\lambda_{\max}^{(k+1)} = \sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}$, which implies that

$$\lambda_{\min}^{(k+1)} = \left(\lambda_{\max}^{(k+1)}\right)^{-1}\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k} \geq \left(\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}\right)^{-1}\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k} = \frac{\sigma y_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}y_k + \sigma y_k^{\mathrm{T}}s_k}.$$

Hence, (23) and (24) hold, which implies that $Q_{k+1}$ is a symmetric positive definite matrix when $\sigma(y_k^{\mathrm{T}}s_k) > 0$. □

From the above theorem, we can easily obtain the following corollary.

**Corollary 1** *Let $Q_{k+1}$ be defined by (6) and $\sigma(y_k^{\mathrm{T}}s_k) > 0$. The spectral condition number of $Q_{k+1}$, $\kappa_2(Q_{k+1})$, is formulated by*

$$\kappa_2(Q_{k+1}) = \left[\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} + \sqrt{\left(\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right)^2 - 4\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}}\right]^2 \Bigg/ 4\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}. \quad (27)$$

*Especially, $\kappa_2(Q_{k+1})$ arrives at the minimum, $(\sqrt{\omega_k} + \sqrt{\omega_k - 1})^2$, when $\sigma = \frac{y_k^{\mathrm{T}}y_k}{s_k^{\mathrm{T}}y_k}$.*

*Proof* According to Theorem 2, (21) and (22) imply that (27) holds. Let

$$t = \left(\omega_k + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right)\Bigg/\left(2\sqrt{\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}}\right), \quad (28)$$

then, according to (27), $\kappa_2(Q_{k+1})$ can be rewritten as follows:

$$\kappa_2(Q_{k+1}) = \psi(t) = \left(t + \sqrt{t^2 - 1}\right)^2, \quad (29)$$

where $\psi(\cdot)$ is a strictly increasing function on $[1, +\infty)$. Note that

$$t \geq \left(2\sqrt{\omega_k}\sqrt{\sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}}\right)\Bigg/\left(2\sqrt{\sigma\frac{s_k^{\mathrm{T}}s_k}{y_k^{\mathrm{T}}s_k}}\right) = \sqrt{\omega_k} \geq 1$$

and the above first inequality takes "=" if and only if $\omega_k = \sigma \frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k}$, namely, $\sigma = \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$. Hence, the minimum of $\kappa_2(Q_{k+1})$ is $(\sqrt{\omega_k} + \sqrt{\omega_k - 1})^2$ when $\sigma = \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$. □

## 2.2 Descent property

For the SPCG method, Theorem 2 shows that the symmetric Perry conjugate gradient directions defined by (5) satisfy the descent property (7), when $\sigma y_k^{\mathrm{T}} s_k > 0$. In fact,

$$d_{k+1}^{\mathrm{T}} g_{k+1} = -g_{k+1}^{\mathrm{T}} Q_{k+1} g_{k+1} \le -\lambda_{\min}^{(k+1)} \|g_{k+1}\|^2 \le -\frac{\sigma y_k^{\mathrm{T}} s_k \|g_{k+1}\|^2}{y_k^{\mathrm{T}} y_k + \sigma y_k^{\mathrm{T}} s_k} < 0. \tag{30}$$

When $\sigma = 1$, $Q_{k+1}$, defined by (6), becomes

$$H_{k+1}^{\mathrm{mBFGS}} = I - \frac{s_k y_k^{\mathrm{T}} + y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \left(1 + \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}\right) \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}. \tag{31}$$

Thus, the method defined by (1) and (5) with $\sigma = 1$ is the famous memoryless BFGS quasi-Newton method [25], denoted by mBFGS.

According to (30), we let $\sigma = c \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$, $c > 0$ and $s_k^{\mathrm{T}} y_k \ne 0$, then it follows from (6) and (30) that

$$Q_{k+1}^{\mathrm{SPD}} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right)\left(I - \frac{y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) + c \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k} \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} \tag{32}$$

and

$$d_{k+1}^{\mathrm{T}} g_{k+1} \le -\frac{c}{1+c} \|g_{k+1}\|_2^2, \tag{33}$$

which shows that the directions defined by (5) with $\sigma = c \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$ satisfy the sufficient descent property (8) for any functions and any line searches. Thus, the method defined by (1) and (5) with $\sigma = c \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$ is called symmetric Perry descent conjugate gradient algorithm, denoted by SPDCG, or by SPDCG(c), to indicate the dependence on the positive constant $c$. Especially, due to Corollary 1, when $c = 1$, the method is called symmetric Perry descent conjugate gradient algorithm with optimal condition number, denoted by SPDOC. The corresponding iteration matrix is denoted by $Q_{k+1}^{\mathrm{SPDOC}}$, i.e.

$$Q_{k+1}^{\mathrm{SPDOC}} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right)\left(I - \frac{y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) + \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k} \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}. \tag{34}$$

In addition, when $\sigma = 0$, then it follows from (6) and (30) that

$$H_{k+1}^{\mathrm{SHS}} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right)\left(I - \frac{y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) \tag{35}$$

and $d_{k+1}^{\mathrm{T}} g_{k+1} \leq 0$. Thus, the method defined by (1) and (5) with $\sigma = 0$ is the symmetric Hestenes-Stiefel method [18], denoted by SHS, which does not satisfy the descent property (7).

## 3 Scaling technology and restarting strategy

According to S.S. Oren and E. Spedicato's idea [20], D.F. Shanno applied the scaling technology to the memoryless BFGS update formula (31) and developed a self-scaling conjugate gradient algorithms [25], i.e., he translated the memoryless BFGS update formula (31) into

$$H_{k+1}(\rho) = \rho\left(I - \frac{s_k y_k^{\mathrm{T}} + y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k} \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) + \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}. \tag{36}$$

Thus, the symmetric Perry matrix $Q_{k+1}$ defined by (6) can be scaled as follows:

$$Q_{k+1}(\rho) = \rho\left(I - \frac{s_k y_k^{\mathrm{T}} + y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k} \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) + \sigma \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}.$$

We substitute $\sigma$ in $Q_{k+1}(\rho)$ with $\rho\sigma$, then $Q_{k+1}(\rho) = \rho Q_{k+1}$, where

$$\rho Q_{k+1} = \rho I - \rho \frac{s_k y_k^{\mathrm{T}} + y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \left(\rho\sigma + \rho \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}\right) \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}. \tag{37}$$

Thus, the line search directions defined by (5) are rewritten as

$$\begin{cases} d_1 = -g_1, \\ d_{k+1} = -\rho Q_{k+1} g_{k+1} \\ \qquad = -\rho g_{k+1} + \rho \dfrac{s_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} y_k + \rho \Big[\frac{y_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} - \big(\sigma + \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}\big) \frac{s_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k}\Big] s_k, \quad \forall k > 1, \end{cases} \tag{38}$$

from which a family of scaling Perry conjugate gradient methods can be deduced.

Based on Beale-Powell restarting strategy [24] (see also [2, 3, 25, 26]), we define the following scheme to compute the directions. When

$$\left|g_{r+1}^{\mathrm{T}} g_r\right| \geq 0.2\|g_{r+1}\|^2 \tag{39}$$

at $r$-th step, we use the directions defined by (38). For $k > r$, the directions $d_{k+1}$ are computed by the following double update scheme:

$$d_{k+1} = -H_{k+1} g_{k+1} \tag{40}$$

with

$$H_{k+1} = H_{r+1} - \frac{s_k y_k^{\mathrm{T}} H_{r+1} + H_{r+1} y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \left(\widetilde{\sigma} + \frac{y_k^{\mathrm{T}} H_{r+1} y_k}{s_k^{\mathrm{T}} y_k}\right) \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} \tag{41}$$

and

$$H_{r+1} = \widehat{\rho} Q_{r+1} = \widehat{\rho}\left[ I - \frac{s_r y_r^{\mathrm{T}} + y_r s_r^{\mathrm{T}}}{s_r^{\mathrm{T}} y_r} + \left( \widehat{\sigma} + \frac{y_r^{\mathrm{T}} y_r}{s_r^{\mathrm{T}} y_r} \right) \frac{s_r s_r^{\mathrm{T}}}{s_r^{\mathrm{T}} y_r} \right], \tag{42}$$

where $\widetilde{\sigma}$, $\widehat{\rho}$ and $\widehat{\sigma}$ are three preset parameters.

Since

$$H_{r+1}^{-1/2} H_{k+1} H_{r+1}^{-1/2} = I - \frac{\widetilde{s}_k \widetilde{y}_k^{\mathrm{T}} + \widetilde{y}_k \widetilde{s}_k^{\mathrm{T}}}{\widetilde{s}_k^{\mathrm{T}} \widetilde{y}_k} + \left( \widetilde{\sigma} + \frac{\widetilde{y}_k^{\mathrm{T}} \widetilde{y}_k}{\widetilde{s}_k^{\mathrm{T}} \widetilde{y}_k} \right) \frac{\widetilde{s}_k \widetilde{s}_k^{\mathrm{T}}}{\widetilde{s}_k^{\mathrm{T}} \widetilde{y}_k}, \tag{43}$$

where $\widetilde{s}_k = H_{r+1}^{-1/2} s_k$, $\widetilde{g}_{k+1} = H_{r+1}^{1/2} g_{k+1}$ and $\widetilde{y}_k = H_{r+1}^{1/2} y_k$, Corollary 1 asserts that $\kappa_2(H_{r+1}^{-1/2} H_{k+1} H_{r+1}^{-1/2})$ arrives at the minimum if $\widetilde{\sigma} = \frac{\widetilde{y}_k^{\mathrm{T}} \widetilde{y}_k}{\widetilde{s}_k^{\mathrm{T}} \widetilde{y}_k} = \frac{y_k^{\mathrm{T}} H_{r+1} y_k}{s_k^{\mathrm{T}} y_k}$. According to (42), Corollary 1 shows that $\kappa_2(H_{r+1})$ is minimal when $\widehat{\sigma} = \frac{y_r^{\mathrm{T}} y_r}{s_r^{\mathrm{T}} y_r}$. We also note that the matrix $H_{r+1}^{-1/2} H_{k+1} H_{r+1}^{-1/2}$ is similar to the matrix $H_{r+1}^{-1} H_{k+1}$, thus

$$\kappa_2(H_{k+1}) \le \kappa_2(H_{r+1}) \kappa_2\big(H_{r+1}^{-1} H_{k+1}\big) = \kappa_2(H_{r+1}) \kappa_2\big(H_{r+1}^{-1/2} H_{k+1} H_{r+1}^{-1/2}\big), \tag{44}$$

which implies that the optimal choices for $\widetilde{\sigma}$ in (41) and $\widehat{\sigma}$ in (42) are

$$\widetilde{\sigma} = \frac{y_k^{\mathrm{T}} H_{r+1} y_k}{s_k^{\mathrm{T}} y_k} \text{ and } \widehat{\sigma} = \frac{y_r^{\mathrm{T}} y_r}{s_r^{\mathrm{T}} y_r}, \tag{45}$$

respectively, such that $\kappa_2(H_{k+1})$ is optimal.

Let $\widehat{g}_{k+1} = H_{r+1} g_{k+1}$ and $\widehat{y}_k = H_{r+1} y_k$, namely,

$$\widehat{g}_{k+1} = \widehat{\rho} g_{k+1} - \widehat{\rho} \frac{s_r^{\mathrm{T}} g_{k+1}}{s_r^{\mathrm{T}} y_r} y_r + \widehat{\rho}\left[ \left( \widehat{\sigma} + \frac{y_r^{\mathrm{T}} y_r}{s_r^{\mathrm{T}} y_r} \right) \frac{s_r^{\mathrm{T}} g_{k+1}}{s_r^{\mathrm{T}} y_r} - \frac{y_r^{\mathrm{T}} g_{k+1}}{s_r^{\mathrm{T}} y_r} \right] s_r \tag{46}$$

and

$$\widehat{y}_k = \widehat{\rho} y_k - \widehat{\rho} \frac{s_r^{\mathrm{T}} y_k}{s_r^{\mathrm{T}} y_r} y_r + \widehat{\rho}\left[ \left( \widehat{\sigma} + \frac{y_r^{\mathrm{T}} y_r}{s_r^{\mathrm{T}} y_r} \right) \frac{s_r^{\mathrm{T}} y_k}{s_r^{\mathrm{T}} y_r} - \frac{y_r^{\mathrm{T}} y_k}{s_r^{\mathrm{T}} y_r} \right] s_r, \tag{47}$$

then the directions $d_{k+1}$ defined by (40) can be reformulated by

$$d_{k+1} = -\widehat{g}_{k+1} + \frac{s_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} \widehat{y}_k - \left[ \left( \widetilde{\sigma} + \frac{\widehat{y}_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k} \right) \frac{s_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} - \frac{\widehat{y}_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} \right] s_k. \tag{48}$$

Hence, we can introduce the following scaling symmetric Perry conjugate gradient method with restarting procedures (SSPCGRP).

**Algorithm 2** (SSPCGRP)

Step 1. Give an initial point $x_1$ and $\varepsilon \ge 0$. Set $k = 1$ and Nrestart $= 0$.

Step 2. Calculate $g_1 = g(x_1)$. If $\|g_1\| \le \varepsilon$, then stop, otherwise, let $d_1 = -g_1$.

Step 3. Calculate steplength $\alpha_k$ using the Wolfe line searches (15) and (16) with initial guess $\alpha_{k,0}$, where $\alpha_{1,0} = 1/\|g_1\|$ and $\alpha_{k,0} = \alpha_{k-1}\|d_{k-1}\|/\|d_k\|$ when $k \ge 2$.

Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$.
Step 5. Calculate $g_{k+1} = g(x_{k+1})$. If $\|g_{k+1}\| \le \varepsilon$ then stop.
Step 6. If the Powell restarting criterion (39) holds, then calculate the directions $d_{k+1}$ via (38) with different $\sigma$ and $\rho$, let $y_r = y_k$ and $s_r = s_k$ (store $y_r$ and $s_r$), set Nrestart = Nrestart + 1 and $k = k + 1$, go to step 3. Otherwise, go to step 7.
Step 7. If Nrestart = 0, then calculate the directions $d_{k+1}$ via (38) with different $\sigma$ and $\rho$, otherwise, calculate $d_{k+1}$ via (48), where $\widehat{y}_k$ and $\widehat{g}_{k+1}$ are computed by (46) and (47), respectively, $\widetilde{\sigma}$, $\widehat{\sigma}$ and $\widehat{\rho}$ are preset parameters.
Step 8. Set $k = k + 1$, go to step 3.

In Algorithm 2, $k$ and Nrestart record the number of iterations and the number of restarting procedures, respectively.

When $\rho = 1$ and $\sigma = c_1 \frac{y_k^T y_k}{s_k^T y_k}$ in (38), and $\widehat{\rho} = 1$, $\widetilde{\sigma} = c_2 \frac{y_k^T H_{r+1} y_k}{s_k^T y_k}$ and $\widehat{\sigma} = c_2 \frac{y_r^T y_r}{s_r^T y_r}$ in (46)–(48), then the SSPCGRP algorithm is denoted by SPDRP, or SPDRP($c_1, c_2$) to indicate the dependence on the positive constants $c_1$ and $c_2$. Especially, when they are equal to 1, i.e., $\sigma = \frac{y_k^T y_k}{s_k^T y_k}$ in (38), $\widetilde{\sigma}$ and $\widehat{\sigma}$ are computed by (45), the condition numbers $\kappa_2(Q_{k+1}) = \kappa_2(\rho Q_{k+1})$, $\kappa_2(H_{k+1})$ and $\kappa_2(H_{r+1})$ are optimal, where $Q_{k+1}$ is defined by (6) with $\sigma = \frac{y_k^T y_k}{s_k^T y_k}$. So, the SSPCGRP algorithm is called the symmetric Perry descent conjugate gradient method with optimal condition numbers and restarting procedures, denoted by SPDOCRP.

When $\rho\sigma = 1$, $\rho = \frac{s_k^T s_k}{y_k^T s_k}$, $\widehat{\rho}\,\widehat{\sigma} = 1$, $\widehat{\rho} = \frac{s_r^T s_r}{y_r^T s_r}$ and $\widetilde{\sigma} = 1$, these formulas (38), (46), (47) and (48) were used by N. Andrei in [2], the SSPCGRP algorithm becomes the SCALCG algorithm with the spectral choice for $\theta_{k+1}$ [2], it is also called Andrei-Perry conjugate gradient method with restarting procedures.

## 4 Convergence

In this section, we analyze the convergence of the symmetric Perry conjugate gradient method (Algorithm 1) and the scaling symmetric Perry conjugate gradient method with restarting procedures (Algorithm 2). For this, we assume that the objective function $f(x)$ satisfies the following assumptions:

H1. $f$ is bounded below in $\mathbb{R}^n$ and $f$ is continuously differentiable in a neighborhood $\mathcal{N}$ of the level set $\mathcal{L} \overset{def}{=} \{x : f(x) \le f(x_0)\}$, where $x_0$ is the starting point of the iteration.
H2. The gradient of $f$ is Lipschitz continuous in $\mathcal{N}$, that is, there exists a constant $L > 0$ such that

$$\left\|\nabla f(\bar{x}) - \nabla f(x)\right\| \le L\|\bar{x} - x\|, \quad \forall \bar{x}, x \in \mathcal{N}. \tag{49}$$

Next, we introduce the spectral condition lemma of the global convergence for an objective function satisfying H1 and H2, which comes from [18], Theorem 4.1.

**Lemma 1** *Let the objective function $f(x)$ satisfy* H1 *and* H2. *Assume that the line search directions of a nonlinear conjugate gradient method satisfy*

$$d_1 = -g_1, d_k = -M_k g_k \quad \forall k > 1, \tag{50}$$

*where $M_k$ is the conjugate gradient iteration matrix, which is a symmetric positive semidefinite matrix. For a nonlinear conjugate gradient method* (1) *and* (50) *satisfying the sufficient descent condition* (8), *if its line search satisfies the Wolfe conditions* (15) *and* (16), *and*

$$\sum_{k=1}^{\infty} (\Lambda_k)^{-2} = +\infty, \tag{51}$$

*where $\Lambda_k$ is the maximum eigenvalue of $M_k$, then* $\liminf_{k\to\infty} \|g_k\| = 0$. *Moreover, if $\Lambda_k \le \widetilde{\Lambda}$ for all $k$, where $\widetilde{\Lambda}$ is a positive constant, then* $\lim_{k\to\infty} \|g_k\| = 0$.

*Remark 2* If $M_k$ is a symmetric positive definite matrix, then the spectral condition (51) can be rewritten as

$$\sum_{k=1}^{\infty} \big(\kappa_2(M_k)\big)^{-2} = +\infty. \tag{52}$$

In fact, by (50), it can be derived that

$$\cos^2\theta_k = \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2 \|g_k\|^2} = \frac{(g_k^{\mathrm{T}} M_k g_k)^2}{g_k^{\mathrm{T}} M_k^{\mathrm{T}} M_k g_k \|g_k\|^2} \ge \frac{\lambda_k^2 \|g_k\|^4}{\Lambda_k^2 \|g_k\|^4} = \big(\kappa_2(M_k)\big)^{-2}, \tag{53}$$

where $\theta_k$ is the angle between $d_k$ and $-g_k$, $\lambda_k$ and $\Lambda_k$ are the minimum eigenvalue and maximum eigenvalue of $M_k$, respectively. The Zoutendijk's condition (Theorem 2.1 of [8]) asserts that (52) implies that the results of Lemma 1 are true.

In what follows, the convergence of these resulting algorithms is proved by evaluating the spectral boundary of the iteration matrix and Lemma 1. The proof method is called the spectral method. It should be pointed out that the proof method also can be applied to the non-symmetric conjugate gradient methods, if the positive square root of the maximum eigenvalue $M_k^{\mathrm{T}} M_k$ substitutes for with the one of $M_k$ in Lemma 1, that is, the maximum singular value of $M_k$ substitutes for the maximum eigenvalue of $M_k$ (see Theorem 3.1 in [17]).

### 4.1 The convergence for uniformly convex functions

Here, we first prove the global convergence of the symmetric Perry conjugate gradient method (SPCG), the scheme (1) and (5) with $Q_{k+1}$ defined by (6), for uniformly convex functions. For this, we introduce the following basic assumption, which is an equivalent condition for a uniformly convex differentiable function.

H3. There exists a constant $m > 0$ such that

$$\big(\nabla f(\bar{x}) - \nabla f(x)\big)^{\mathrm{T}} (\bar{x} - x) \ge m\|\bar{x} - x\|^2 \quad \forall \bar{x},\ x \in \mathcal{N}. \tag{54}$$

**Theorem 3** *Assume that* H1, H2 *and* H3 *hold. Let* $\nu_0$ *and* $\nu_1$ *be two positive constants. For the symmetric Perry conjugate gradient method* (1) *and* (5) *with* $\nu_0 \leq \sigma \leq \nu_1$, *the Wolfe line searches* (15) *and* (16) *are implemented. If* $g_1 \neq 0$ *and steplength* $\alpha_k > 0$ *for* $k \geq 1$, *then* $g_k = 0$ *for some* $k > 1$, *or* $\lim_{k\to\infty} \|g_k\| = 0$.

*Proof* Assume that $g_k \neq 0$, $\forall k \in \mathbb{N}$. Below, by induction, we first prove that the line search direction $d_k$, defined by (5), satisfies the sufficient descent property (8).

When $k = 1$, $d_1^{\mathrm{T}} g_1 = -\|g_1\|^2 < 0$. From (16), it follows that $s_1^{\mathrm{T}} y_1 \geq -(1 - b_2)\alpha_1 d_1^{\mathrm{T}} g_1 > 0$.

Now, assume that $d_k^{\mathrm{T}} g_k \leq -\frac{\nu_0 m}{L^2+\nu_0 m}\|g_k\|^2$. Then, it follows from (16) that $s_k^{\mathrm{T}} y_k \geq -(1-b_2)\alpha_k d_k^{\mathrm{T}} g_k > 0$. So, (30) and the assumptions H2 and H3 imply that

$$d_{k+1}^{\mathrm{T}} g_{k+1} \leq -\frac{\sigma y_k^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} y_k + \sigma y_k^{\mathrm{T}} s_k}\|g_{k+1}\|^2 \leq -\frac{\nu_0 m}{L^2 + \nu_0 m}\|g_{k+1}\|^2.$$

Hence, by induction, the sufficient descent property (8) holds.

Next, we prove that $\lambda_{\max}^{(k+1)}$, the maximum eigenvalue of $Q_{k+1}$ defined by (6), is uniformly bounded above. From the above analysis, it can be derived that $s_k^{\mathrm{T}} y_k > 0$. So, from (23) in Theorem 2, it can be deduced that

$$\lambda_{\max}^{(k+1)} \leq \frac{y_k^{\mathrm{T}} y_k s_k^{\mathrm{T}} s_k}{(s_k^{\mathrm{T}} y_k)^2} + \sigma\frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k} \leq \frac{L^2\|s_k\|^4}{m^2\|s_k\|^4} + \frac{\sigma\|s_k\|^2}{m\|s_k\|^2} = \frac{L^2}{m^2} + \frac{\nu_1}{m}. \tag{55}$$

Therefore, Lemma 1 claims that $\lim_{k\to\infty}\|g_k\| = 0$. □

*Remark 3* Theorem 3 shows that the memoryless BFGS quasi-Newton method and the method SPDCG are convergent for uniformly convex functions under the Wolfe line searches. In fact, the global convergence of the method SPDCG and the method mBFGS results from the following inequalities

$$m \leq \frac{y_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} s_k} \leq \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k} \leq \frac{L^2}{m}. \tag{56}$$

Next, we prove the global convergence of the SSPCGRP method for uniformly convex functions.

**Theorem 4** *Assume that* H1, H2 *and* H3 *hold, and that* $\nu_0$ *and* $\nu_1$ *are two positive constants. Let the sequence* $\{x_k\}$ *be generated by the SSPCGRP algorithm* (*Algorithm* 2), *where the five different parameters* $\sigma$, $\rho$, $\widetilde{\sigma}$, $\widehat{\sigma}$ *and* $\widehat{\rho}$ *satisfy* $\nu_0 \leq \sigma, \rho, \widetilde{\sigma}, \widehat{\sigma}, \widehat{\rho} \leq \nu_1$. *If* $g_1 \neq 0$, *and steplength* $\alpha_k > 0$ *for* $k \geq 1$, *then* $g_k = 0$ *for some* $k > 1$, *or* $\lim_{k\to\infty}\|g_k\| = 0$.

*Proof* First, we note that $\widetilde{y}_k^{\mathrm{T}}\widetilde{s}_k = y_k^{\mathrm{T}} s_k$ for all $k \geq 1$. If $\widetilde{y}_k^{\mathrm{T}}\widetilde{s}_k > 0$, we can denote the minimum and the maximum eigenvalues of the matrix $H_{r+1}^{-1/2} H_{k+1} H_{r+1}^{-1/2}$ by $\widetilde{\lambda}_{\min}^{(k+1)}$ and $\widetilde{\lambda}_{\max}^{(k+1)}$, respectively. We also denote the minimum and the maximum eigenvalues of the matrix $H_{r+1}$ by $\widehat{\lambda}_{\min}^{(r+1)}$ and $\widehat{\lambda}_{\max}^{(r+1)}$, respectively. Thus, (42), (43), Theorem 2 and the assumptions H2 and H3 imply that

$$\widetilde{\lambda}_{\max}^{(k+1)} \le \widetilde{\omega}_k + \widetilde{\sigma}\frac{\widetilde{s}_k^{\mathrm{T}}\widetilde{s}_k}{\widetilde{s}_k^{\mathrm{T}}\widetilde{y}_k}, \widetilde{\lambda}_{\min}^{(k+1)} \ge \frac{\widetilde{\sigma}\widetilde{y}_k^{\mathrm{T}}\widetilde{s}_k}{\widetilde{y}_k^{\mathrm{T}}\widetilde{y}_k + \widetilde{\sigma}\widetilde{y}_k^{\mathrm{T}}\widetilde{s}_k} \ge \frac{\nu_0 y_k^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} H_{r+1} y_k + \nu_0 y_k^{\mathrm{T}} s_k} \tag{57}$$

and

$$\widehat{\lambda}_{\max}^{(r+1)} \le \widehat{\rho}\omega_r + \widehat{\rho}\,\widehat{\sigma}\frac{s_r^{\mathrm{T}} s_r}{s_r^{\mathrm{T}} y_r} \le \nu_1\frac{L^2 + \nu_1 m}{m^2}, \widehat{\lambda}_{\min}^{(r+1)} \ge \frac{\widehat{\rho}\,\widehat{\sigma} y_r^{\mathrm{T}} s_r}{y_r^{\mathrm{T}} y_r + \widehat{\sigma} y_r^{\mathrm{T}} s_r} \ge \frac{\nu_0^2 m}{L^2 + \nu_0 m}, \tag{58}$$

where $\widetilde{\omega}_k = \frac{\widetilde{y}_k^{\mathrm{T}}\widetilde{y}_k\widetilde{s}_k^{\mathrm{T}}\widetilde{s}_k}{(\widetilde{s}_k^{\mathrm{T}}\widetilde{y}_k)^2}$ and $\omega_r = \frac{y_r^{\mathrm{T}} y_r s_r^{\mathrm{T}} s_r}{(s_r^{\mathrm{T}} y_r)^2}$.

In what follows, by induction, we prove that $\widetilde{y}_k^{\mathrm{T}}\widetilde{s}_k > 0$ and the sufficient descent property (8) is true for all $k$.

If the Powell restarting criterion (39) never holds for all $k \ge 1$, the iteration matrix is $\rho Q_{k+1}$. Thus, similar to Theorem 3, it can be easily shown that the results of Theorem 4 are true.

Suppose that $k_0$ is the first natural number such that the Powell restarting criterion (39) is true, then Nrestart $\ge 1$. Similar to Theorem 3, it can be obtained that for $k = 1, 2, \ldots, k_0$, $\widetilde{y}_k^{\mathrm{T}}\widetilde{s}_k = y_k^{\mathrm{T}} s_k > 0$ and

$$d_{k+1}^{\mathrm{T}} g_{k+1} \le -\frac{\rho\sigma y_k^{\mathrm{T}} s_k \|g_{k+1}\|^2}{y_k^{\mathrm{T}} y_k + \sigma y_k^{\mathrm{T}} s_k} \le -\frac{\rho\sigma m\|g_{k+1}\|^2}{L^2 + \sigma m} \le -\frac{\nu_0^2 m}{L^2 + \nu_0 m}\|g_{k+1}\|^2.$$

So, it follows from (16) and the above inequality that

$$s_{k+1}^{\mathrm{T}} y_{k+1} \ge -(1 - b_2)\alpha_{k+1} d_{k+1}^{\mathrm{T}} g_{k+1} > 0.$$

If the Powell restarting criterion (39) holds for $k + 2$, $d_{k+2}$ is calculated by (38). Thus, Theorem 2 and the assumptions H2 and H3 claim that

$$d_{k+2}^{\mathrm{T}} g_{k+2} \le -\frac{\rho\sigma y_{k+1}^{\mathrm{T}} s_{k+1}\|g_{k+2}\|^2}{y_{k+1}^{\mathrm{T}} y_{k+1} + \sigma y_{k+1}^{\mathrm{T}} s_{k+1}} \le -\frac{\rho\sigma m\|g_{k+2}\|^2}{L^2 + \sigma m} \le -\frac{\nu_0^2 m}{L^2 + \nu_0 m}\|g_{k+1}\|^2.$$

If the Powell restarting criterion (39) does not hold for $k + 2$, $d_{k+2}$ is calculated by (48) with (46) and (47) (i.e., (40)–(42)). Since Nrestart $\ge 1$, from (40)–(42) and Theorem 2, it can be obtained that

$$\begin{aligned} d_{k+2}^{\mathrm{T}} g_{k+2} &= -g_{k+2}^{\mathrm{T}} H_{k+2} g_{k+2} = -\widetilde{g}_{k+2}^{\mathrm{T}} H_{r+1}^{-1/2} H_{k+2} H_{r+1}^{-1/2}\widetilde{g}_{k+2} \\ &\le -\widetilde{\lambda}_{\min}^{(k+2)}\widetilde{g}_{k+2}^{\mathrm{T}}\widetilde{g}_{k+2} = -\widetilde{\lambda}_{\min}^{(k+2)} g_{k+2}^{\mathrm{T}} H_{r+1} g_{k+2} \le -\widetilde{\lambda}_{\min}^{(k+2)}\widehat{\lambda}_{\min}^{(r+1)}\|g_{k+2}\|^2. \end{aligned}$$

By (57), (58), the assumptions H2 and H3, it is yielded that

$$\widetilde{\lambda}_{\min}^{(k+2)} \ge \frac{\nu_0 y_{k+1}^{\mathrm{T}} s_{k+1}}{\widehat{\lambda}_{\max}^{(r+1)}\|y_{k+1}\|^2 + \nu_0 y_{k+1}^{\mathrm{T}} s_{k+1}} \ge \frac{\nu_0 m^3}{(L^2 + \nu_1 m)\nu_1 L^2 + \nu_0 m^3}.$$

Therefore,

$$d_{k+2}^{\mathrm{T}} g_{k+2} \leq -\frac{\nu_0^3 m^4 \|g_{k+2}\|^2}{(L^2+\nu_0 m)((L^2+\nu_1 m)\nu_1 L^2+\nu_0 m^3)}$$

which, together with (16), implies that

$$s_{k+2}^{\mathrm{T}} y_{k+2} \geq -(1-b_2)\alpha_{k+2} d_{k+2}^{\mathrm{T}} g_{k+2} > 0.$$

By induction, it follows that $s_k^{\mathrm{T}} y_k > 0$ for all $k$ and the directions generated by the SSPCGRP algorithm (Algorithm 2) satisfy the sufficient descent property (8) with

$$c_0 = \min\left\{\frac{\nu_0^2 m}{L^2+\nu_0 m}, \frac{\nu_0^3 m^4}{(L^2+\nu_0 m)((L^2+\nu_1 m)\nu_1 L^2+\nu_0 m^3)}\right\}.$$

Next, we prove that $\kappa_2(H_{k+1})$ is bounded above. Since $\widetilde{s}_k = H_{r+1}^{-1/2} s_k$ and $\widetilde{y}_k = H_{r+1}^{1/2} y_k$, from (57), (58) and the assumptions H2 and H3, it can be derived that

$$\widetilde{\omega}_k = \frac{\widetilde{y}_k^{\mathrm{T}} \widetilde{y}_k \widetilde{s}_k^{\mathrm{T}} \widetilde{s}_k}{(\widetilde{s}_k^{\mathrm{T}} \widetilde{y}_k)^2} = \frac{y_k^{\mathrm{T}} H_{r+1} y_k s_k^{\mathrm{T}} H_{r+1}^{-1} s_k}{(s_k^{\mathrm{T}} y_k)^2} \leq \frac{\widehat{\lambda}_{\max}^{(r+1)} \omega_k}{\widehat{\lambda}_{\min}^{(r+1)}} \leq \frac{(L^2+\nu_0 m)(L^2+\nu_1 m)}{\nu_0 m^3} \frac{L^2}{m^2},$$

$$\widetilde{\sigma} \frac{\widetilde{s}_k^{\mathrm{T}} \widetilde{s}_k}{\widetilde{s}_k^{\mathrm{T}} \widetilde{y}_k} = \widetilde{\sigma} \frac{s_k^{\mathrm{T}} H_{r+1}^{-1} s_k}{s_k^{\mathrm{T}} y_k} \leq \widetilde{\sigma} \frac{s_k^{\mathrm{T}} s_k}{\widehat{\lambda}_{\min}^{(r+1)} s_k^{\mathrm{T}} y_k} \leq \nu_1 \frac{L^2+\nu_0 m}{\nu_0^2 m^2},$$

$$\widetilde{\lambda}_{\max}^{(k+1)} \leq \frac{(L^2+\nu_0 m)(L^2+\nu_1 m)}{\nu_0 m^3} \frac{L^2}{m^2} + \nu_1 \frac{L^2+\nu_0 m}{\nu_0^2 m^2}$$

and

$$\widetilde{\lambda}_{\min}^{(k+1)} \geq \frac{\widetilde{\sigma} y_k^{\mathrm{T}} s_k}{\widehat{\lambda}_{\max}^{(r+1)} y_k^{\mathrm{T}} y_k + \widetilde{\sigma} y_k^{\mathrm{T}} s_k} \geq \frac{\nu_0 m^3}{(L^2+\nu_1 m)\nu_1 L^2+\nu_0 m^3}.$$

Thus, it follows from (44), (58) and above two inequalities that

$$\kappa_2(H_{k+1}) \leq \frac{\widehat{\lambda}_{\max}^{(r+1)} \widetilde{\lambda}_{\max}^{(k+1)}}{\widehat{\lambda}_{\min}^{(r+1)} \widetilde{\lambda}_{\min}^{(k+1)}} \leq \left(L^2+\nu_1 m\right)^3 \frac{((L^2+\nu_1 m)\nu_1 L^2+\nu_1 m^3)^2}{\nu_0^5 m^{11}}.$$

If $d_{k+1}$ is calculated by (38), the iteration matrix is defined by (37). So, Corollary 1, (28), (29) and the assumptions H2 and H3 imply that

$$t = \left(\omega_k + \sigma \frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k}\right) \Bigg/ \left(2\sqrt{\sigma \frac{s_k^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} s_k}}\right) \leq \sqrt{\frac{m}{\sigma}} \frac{L^2+\sigma m}{2m^2} \leq \sqrt{\frac{m}{\nu_0}} \frac{L^2+\nu_1 m}{2m^2}$$

and

$$\kappa_2(\rho Q_{k+1}) = \kappa_2(Q_{k+1}) = \psi(t) \leq \psi\left(\sqrt{\frac{m}{\nu_0}} \frac{L^2+\nu_1 m}{2m^2}\right),$$

where $\psi(\cdot)$ is defined by (29).

Hence, the spectral condition number of the iteration matrix of Algorithm 2 is uniformly bounded above, which claims that the results of Theorem 4 are true according to Remark 2. □

From (56) and this theorem, it can be shown that the SPDRP algorithm and the SCALCG algorithm with the spectral choice [2] are global convergence for uniformly convex functions under the Wolfe line searches.

## 4.2 The convergence for general nonlinear functions

For general nonlinear functions, we first have following result for the symmetric Perry conjugate gradient method.

**Theorem 5** *Assume that* H1 *and* H2 *hold. For the symmetric Perry conjugate gradient method* (1) *and* (5) *with* $\sigma = c\frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$, *where $c$ is a positive constant, if the line searches satisfy the Wolfe conditions* (15) *and* (16), *then* $\lim_{k\to\infty} \|y_k\| = 0$ *implies that* $\liminf_{k\to\infty} \|g_k\| = 0$.

*Proof* Denote the maximum eigenvalue of the iteration matrix $Q_{k+1}$ by $\lambda_{\max}^{(k+1)}$. The Wolfe condition (16) leads to $s_k^{\mathrm{T}} y_k \geq -(1-b_2)s_k^{\mathrm{T}} g_k$, which, together with (33), implies that

$$\omega_k \leq \frac{y_k^{\mathrm{T}} y_k s_k^{\mathrm{T}} s_k}{(1-b_2)^2(-s_k^{\mathrm{T}} g_k)^2} = \frac{y_k^{\mathrm{T}} y_k g_k Q_k^2 g_k}{(1-b_2)^2(g_k P_k g_k)(-d_k^{\mathrm{T}} g_k)} \leq \frac{(1+c)\lambda_{\max}^{(k)}\|y_k\|^2}{(1-b_2)^2 c\|g_k\|^2}. \tag{59}$$

Thus,

$$\lambda_{\max}^{(k+1)} \leq \omega_k + \sigma\frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k} = (1+c)\omega_k \leq \frac{(1+c)^2\lambda_{\max}^{(k)}\|y_k\|^2}{(1-b_2)^2 c\|g_k\|^2} = \lambda_{\max}^{(k)} c_5^2\frac{\|y_k\|^2}{\|g_k\|^2},$$

where $c_5 = \frac{1+c}{(1-b_2)\sqrt{c}}$. So,

$$\lambda_{\max}^{(k+1)} \leq \lambda_{\max}^{(k-1)} c_5^4 \frac{\|y_{k-1}\|^2}{\|g_{k-1}\|^2}\frac{\|y_k\|^2}{\|g_k\|^2} \leq \cdots \leq \lambda_{\max}^{(1)} c_5^{2k}\prod_{j=1}^{k}\frac{\|y_j\|^2}{\|g_j\|^2}. \tag{60}$$

Now assume that $\lim_{k\to\infty}\|y_k\| = 0$, $\liminf_{k\to\infty}\|g_k\| = \varepsilon > 0$. Then there exists a positive integer $N_0$ such that for $j > N_0$, $\frac{\|y_j\|}{\|g_j\|} \leq c_5^{-1}$. Let $C_N = \lambda_{\max}^{(1)}\prod_{j=1}^{N_0} c_5^2\frac{\|y_j\|^2}{\|g_j\|^2}$, thus,

$$\lambda_{\max}^{(k+1)} \leq \lambda_{\max}^{(1)}\prod_{j=1}^{N_0} c_5^2\frac{\|y_j\|^2}{\|g_j\|^2}\prod_{j=N_0+1}^{k} c_5^2\frac{\|y_j\|^2}{\|g_j\|^2} \leq C_N.$$

Therefore, Lemma 1 and (33) claim that $\lim_{k\to\infty}\|g_k\| = 0$, which contradicts the above assumption. So, $\lim_{k\to\infty}\|y_k\| = 0$ implies that $\liminf_{k\to\infty}\|g_k\| = 0$. □

Next, we prove the global convergence of the SSPCGRP algorithm (Algorithm 2) for general nonlinear functions.

**Theorem 6** *Assume that* H1 *and* H2 *hold. Let the sequence* $\{x_k\}$ *be generated by the SSPCGRP algorithm with* $\sigma = c\frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$ *and* $\nu_0 \le \rho \le \nu_1$ *in* (38), *where* $c$, $\nu_0$ *and* $\nu_1$ *are positive constants. If the line searches satisfy the Wolfe conditions* (15) *and* (16), *then* $\lim_{k\to\infty} \|y_k\| = 0$ *implies that* $\liminf_{k\to\infty} \|g_k\| = 0$.

*Proof* If $\liminf_{k\to\infty} \|g_k\| \ne 0$ as $\|y_k\| \to 0$, then, for some $\varepsilon > 0$, there exists a positive integer $N_1$ such that $\|g_{k+1}\| > \varepsilon$ and $\|y_k\| \le 0.8\varepsilon$ as $k \ge N_1$. Thus,

$$\begin{aligned}\|g_{k+1}\|^2 &= g_{k+1}^{\mathrm{T}} g_k + g_{k+1}^{\mathrm{T}} y_k \le g_{k+1}^{\mathrm{T}} g_k + \|g_{k+1}\| \|y_k\| \\ &\le g_{k+1}^{\mathrm{T}} g_k + \|g_{k+1}\| 0.8\varepsilon \le g_{k+1}^{\mathrm{T}} g_k + 0.8 \|g_{k+1}\|^2.\end{aligned}$$

So, $g_{k+1}^{\mathrm{T}} g_k \ge 0.2\|g_{k+1}\|^2$ for $k \ge N_1$, which means that the directions $d_{k+1}$ are calculated by (38) for $k \ge N_1$, that is,

$$d_{k+1} = -\rho Q_{k+1} g_{k+1},$$

where $Q_{k+1}$ is defined by (6). Since $\sigma = c\frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$ and $\nu_0 \le \rho \le \nu_1$, from Theorem 2, it follows that

$$d_{k+1}^{\mathrm{T}} g_{k+1} = -\rho g_{k+1}^{\mathrm{T}} Q_{k+1} g_{k+1} \le -\frac{\rho\sigma y_k^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} y_k + \sigma y_k^{\mathrm{T}} s_k} \|g_{k+1}\|^2 \le -\frac{c\nu_0}{1+c} \|g_{k+1}\|^2. \tag{61}$$

For convenience, we also denote the maximum eigenvalue of $\rho Q_{k+1}$ by $\lambda_{\max}^{(k+1)}$. Analogous to (60), it can be derived that

$$\lambda_{\max}^{(k+1)} \le \rho\left(\omega_k + \sigma\frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k}\right) \le \nu_1 c_5^2 \frac{\|y_k\|^2}{\|g_k\|^2} \lambda_{\max}^{(k)} \le \cdots \le \lambda_{\max}^{(N_1)} \prod_{j=N_1}^{k} \nu_1 c_5^2 \frac{\|y_j\|^2}{\|g_j\|^2},$$

where $c_5 = \frac{1+c}{(1-b_2)\sqrt{c}}$. We substitute $\lambda_{\max}^{(N_1)}$ and $N_1$ for $\lambda_{\max}^{(1)}$ and 1 in (60), respectively, then, similar to Theorem 5, it can be obtain from Lemma 1 and (61) that $\lim_{k\to\infty} \|y_k\| = 0$ implies that $\liminf_{k\to\infty} \|g_k\| = 0$. □

The above two theorems show that the SPDCG($c$) algorithm and the SPDRP($c_1, c_2$) algorithm are global convergence for the nonconvex functions under the Wolfe line searches, as $\lim_{k\to\infty} \|y_k\| = 0$. The condition for the global convergence, $\lim_{k\to\infty} \|y_k\| = 0$, was used by J.Y. Han, et al. in [14].

## 5 Numerical experiments

In this section, we demonstrate our algorithms: SPDCG and SPDRP, and compare them with the CG_DESCENT algorithm [12], the SCALCG algorithm with the spectral choice [2], the mBFGS algorithm (a special form of the SPCG algorithm with

$\sigma = 1$) and the RSPDCGs algorithm [19] whose line search directions are formulated by

$$d_1 = -g_1, \qquad d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k \geq 1,$$

where

$$\beta_k = \frac{1}{\eta_k^d}\left(y_k - \frac{\|y_k\|^2}{\eta_k^d} d_k\right)^{\mathrm{T}} g_{k+1} \quad \text{with } \eta_k^d = \begin{cases} y_k^{\mathrm{T}} d_k, & \text{if } \|g_k\|^2 \geq \eta\alpha_k \|d_k\|^2, \\ \alpha_k \|d_k\|^2, & \text{otherwise.} \end{cases} \tag{62}$$

In numerical experiments, we let $\eta = 10^{-5}$. The RSPDCGs algorithm is fully detailed in [19].

The numerical experiments use two groups test functions, one group (145 test functions) is taken from the CUTEr [9] library, referring to website:

http://www.cuter.rl.ac.uk/,

which is only used to test mBFGS, SPDCG, RSPDCGs and CG_DESCENT algorithms. In order to compare with the SCALCG algorithm, the second group consists of the 73 unconstrained problems but the 71-st in SCALCG Fortran software package coded by N. Andrei, referring to website:

http://camo.ici.ro/forum/SCALCG/.

For the second group, each test function is made ten experiments with the number of variable $1000, 2000, \ldots, 10000$, respectively. The starting points used are those given in the code, SCALCG.

The SPDCG, mBFGS and RSPDCGs algorithms are coded according to the package, CG_DESCENT (C language, Version 5.3), with minor revisions and implement the approximate Wolfe line searches with the default parameters in CG_DESCENT [10, 12]. The package, CG_ DESCENT, can be got from Hager's web page at

http://www.math.ufl.edu/~hager/.

In addition, in order to compare with the SCALCG algorithm, all subroutines of the SPDRP algorithm are written in Fortran 77 with the double precision, and the SPDRP algorithm uses the Wolfe line searches in the SCALCG Fortran code.

The termination criterion of all algorithms is that $\|g\|_\infty < 10^{-6}$, where $\|\cdot\|_\infty$ is the infinity norm of a vector. The maximum number of iterations is $500n$, where $n$ is the number of variables. The tests are performed on PC (Dell Inspiron 530), Intel® Core™ 2 Duo, E4600, 2.40 GHz, 2.39 GHz, RAM 2.00 GB, with the gcc and g77 compilers.

The SPDCG algorithm is a special form of the SPCG algorithm (Algorithm 1) with $\sigma = c\frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$ in (5) (see also step 6 in Algorithm 1). Thus the line search directions are formulated by

$$d_1 = -g_1, \qquad d_{k+1} = -g_{k+1} + \beta_k d_k + \gamma_k y_k, \quad \forall k \geq 1 \tag{63}$$

with $\beta_k = \frac{y_k^{\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k} - (1+c)\frac{y_k^{\mathrm{T}} y_k}{d_k^{\mathrm{T}} y_k}\frac{d_k^{\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k}$, $\gamma_k = \frac{d_k^{\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k}$, and the iteration matrix is defined by (32). For the SPDCG algorithm, we test several different values of $c$ in $\beta_k$ of (63)

**Fig. 1** Performance based on Nite of SPDOC, mBFGS, CG_DESCENT and RSPDCGs for large scale problems

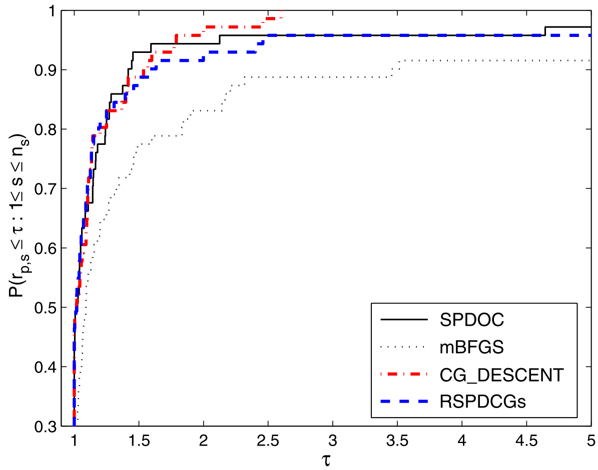
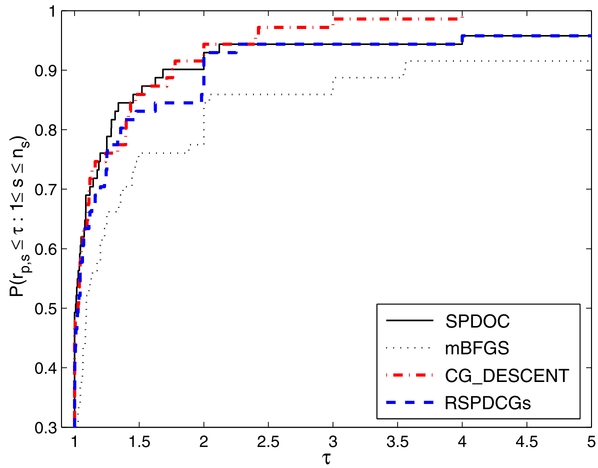



**Fig. 2** Performance based on CPU time of SPDOC, mBFGS, CG_DESCENT and RSPDCGs for large scale problems




on the first group of test functions and find that the performance [6] is slightly better when $c = 1$ (SPDOC algorithm) than that when $c$ is taken other values.

For the first group of test functions, to compare the algorithms: mBFGS and SPDOC with the RSPDCGs and CG_DESCENT algorithms. we divide the group into two parts: large scale problems, whose numbers of variables are not less than 100 (72 test functions), and small scale problems, whose numbers of variables are less than 100 (73 test functions).

Figures 1 and 2 present that their Dolan-Moré performance profiles for large scale problems based on Nite (the number of iterations) and CPU time, respectively. Figures 3 and 4 present the Dolan and Moré performance profiles of these algorithms for small scale problems with relative to Nite and CPU time, respectively. Figures 1 and 2 show that for large scale problems, the performance of the SPDOC algorithm is similar to that of the CG_DESCENT algorithm and their performances are better than that of the others; the performance of the RSPDCGs algorithm is better than

**Fig. 3** Performance based on Nite of SPDOC, mBFGS, CG_DESCENT and RSPDCGs for small scale problems

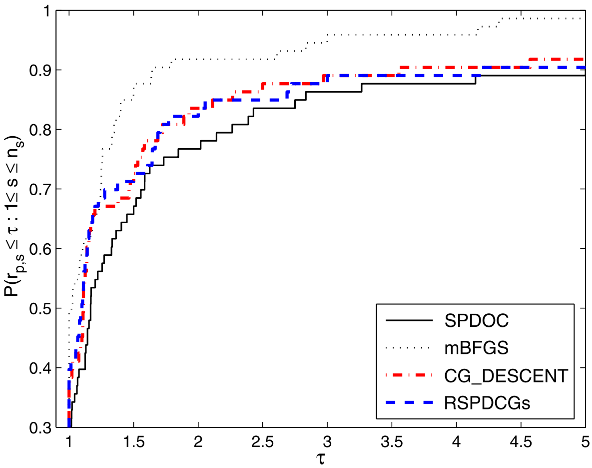
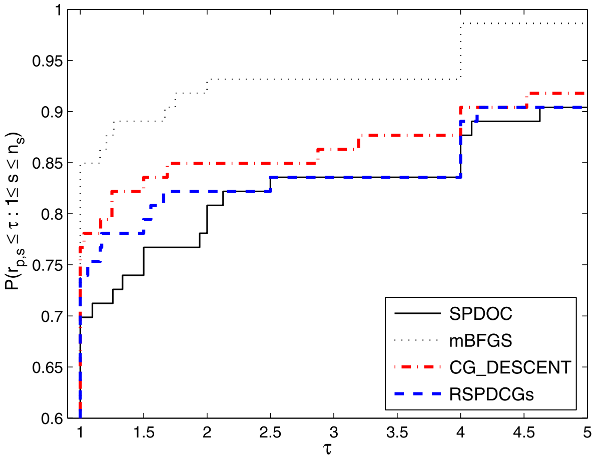



**Fig. 4** Performance based on CPU time of SPDOC, mBFGS, CG_DESCENT and RSPDCGs for small scale problems




that of the mBFGS algorithm. For small scale problems, Figs. 3 and 4 show that the mBFGS algorithm is best, which means that SPDOC and CG_DESCENT algorithms are more suitable for solving large scale problems. It should be pointed out that although mBFGS algorithm give better results on the small problems, it fails to solve three problems of the CUTEr test set. In the recent paper [13], it is observed that for the small ill-conditioned quadratic PALMER test problems in CUTEr, the gradients generated by the conjugate gradient method quickly lose orthogonality due to numerical errors. See Hager and Zhang's paper [13] for a strategy for handling these ill-conditioned problems.

The SPDRP algorithm is a special case of the SSPCGRP algorithm (Algorithm 2) with $\rho = 1$ and $\sigma = c_1 \frac{y_k^T y_k}{s_k^T y_k}$ in (38), and $\widehat{\rho} = 1$, $\widetilde{\sigma} = c_2 \frac{y_k^T H_{r+1} y_k}{s_k^T y_k}$ and $\widehat{\sigma} = c_2 \frac{y_r^T y_r}{s_r^T y_r}$ in (46)–(48). For the SPDRP algorithm, we also test several different values of $c$ on the second group of test functions, we find the performance is slightly better when $c_1 = c_2 = 1$ (i.e., the SPDOCRP algorithm) than that when $c_1$ and $c_2$ are taken other

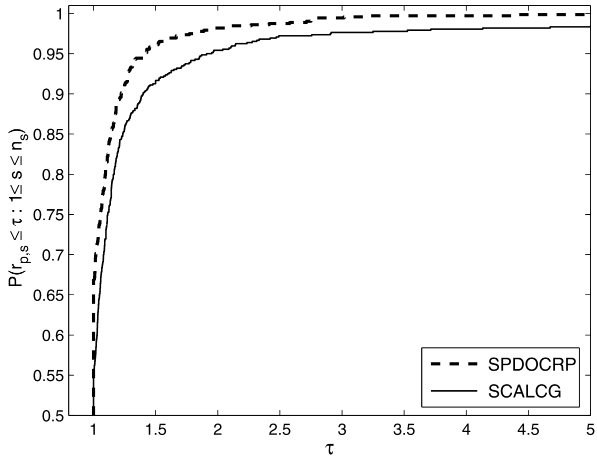



**Fig. 5** Performance based on Nite of SPDOCRP and SCALCG algorithms for the second group of test functions

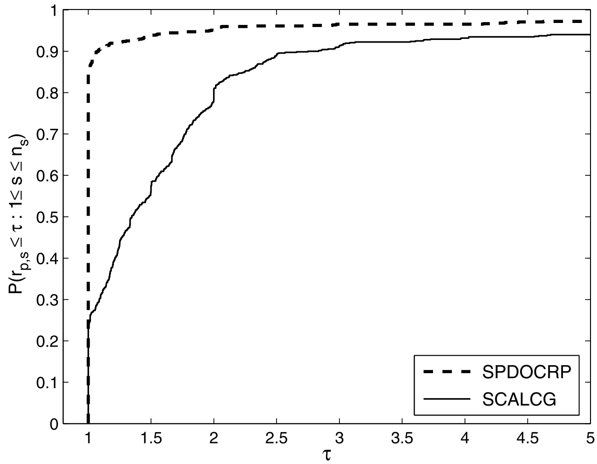



**Fig. 6** Performance based on CPU time of SPDOCRP and SCALCG algorithms for the second group of test functions

values. Next, we compare the SPDOCRP algorithm with the SCALCG algorithms, using the second group of test functions. Figures 5 and 6 present that their Dolan-Moré performance profiles based on Nite and CPU time, respectively. The SPDOCRP algorithm and the SCALCG algorithm use the restarting strategy and the double update scheme, but the SPDOCRP algorithm has the optimal spectral condition number of the iteration matrix, so the SPDOCRP algorithm displays better numerical performance than the SCALCG algorithm.

So, the preliminary numerical experiments show that SPDOC and SPDOCRP are very effective algorithms for the large scale unconstrained optimization problems.

In addition, for the SPDCG algorithm, the inequality (33) shows that the descent degree of the line search directions of the algorithm becomes higher and higher as the value of $c$ increases, but the performance of the algorithm is not directly proportional to $c$. In fact, the line search directions generated by the SPDCG algorithm vary with the value of $c$. What kind of criterion can be used to evaluate the performance of an algorithm? Does the criterion exist? These are still open problems. Of course, the condition number and the descent property are two important factors.

In the end, it should be pointed out that the version 5.3 of CG_ DESCENT (C code) uses a new formula for $\beta_k$,

$$\beta_k = \frac{1}{y_k^{\mathrm{T}} d_k}\left(y_k - \frac{\|y_k\|^2}{y_k^{\mathrm{T}} d_k} d_k\right)^{\mathrm{T}} g_{k+1}, \tag{64}$$

i.e., $\eta \equiv 0$ in (62), instead of

$$\beta_k = \frac{1}{y_k^{\mathrm{T}} d_k}\left(y_k - \frac{2\|y_k\|^2}{y_k^{\mathrm{T}} d_k} d_k\right)^{\mathrm{T}} g_{k+1},$$

presented in [10]. In [19], we proved that $\beta_k$ formulated by (64) makes the spectral condition number of the iteration matrix defined by (4) with $u = s_k$ optimal.

## 6 Conclusion

In [18], we presented a rank one updating formula for the iteration matrix of the conjugate gradient methods:

$$M_{k+1} = M_{k+1}^{shs} + \frac{s_k \xi_k^{\mathrm{T}}}{y_k^{\mathrm{T}} \xi_k}, \quad \forall \xi_k \in \mathbb{R}^n,$$

where the symmetric Hestenes-Stiefel matrix $M_{k+1}^{shs}$ is defined by

$$M_{k+1}^{shs} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right)\left(I - \frac{y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right).$$

If we replace $D_{k+1}$ with $M_{k+1}^{shs}$ in (11), the symmetric Perry matrix (6) can be rewritten as

$$Q_{k+1} = M_{k+1}^{shs} + \sigma \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k},$$

that is, if we apply the Powell symmetric technique to $D_{k+1}$ in (11), we can obtain $M_{k+1}^{shs}$, then we add a rank one update $\sigma \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}$ to $M_{k+1}^{shs}$, we can also deduce the symmetric Perry matrix (6).

For the parameter $\sigma$ in SPCG algorithm, besides the cases mentioned above, there also exist other choices, such as $\sigma = c^2 \frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k}$, $\sigma = c \frac{s_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} s_k}$, $\sigma = c \frac{s_k^{\mathrm{T}} y_k}{y_k^{\mathrm{T}} y_k}$, $\sigma = c \frac{s_k^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} y_k}$, $\sigma = c \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} s_k}$, and so on, where $c > 0$.

For the SSPCGRP algorithm, when $\rho\sigma = 1$, $\sigma = \frac{y_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} y_k}$, $\widehat{\sigma} = \frac{y_r^{\mathrm{T}} y_r}{s_r^{\mathrm{T}} y_r}$, $\widehat{\rho}\,\widehat{\sigma} = \widetilde{\sigma} = 1$, these formulas (38), (46), (47) and (48) were suggested by D. F. Shanno in [25] and [26]. When $\rho = \sigma = \widehat{\rho} = \widetilde{\sigma} = \widehat{\sigma} = 1$, the SSPCGRP algorithm becomes the memoryless BFGS conjugate gradient method with restarting procedures. Therefore, it is worthy of studying further how the parameters $\sigma$ and $\rho$ are chosen to construct more effective nonlinear conjugate gradient algorithms.

The condition number of $Q_{k+1}$ defined by (6) only depends on the parameter $\sigma$ and the condition number of $\rho Q_{k+1}$ is the same as the one of $Q_{k+1}$ (see (37)), so, we let $\rho = 1$ and $\widehat{\rho} = 1$ in SSPCGRP algorithm. That is to say, $\sigma$ can scale the symmetric Perry iteration matrix $Q_{k+1}$. Therefore, the symmetric Perry conjugate gradient methods have the self-scaling property, Similarly, $\sigma$ can also alter the maximum and minimum eigenvalues of the Perry iteration matrix $P_{k+1}$ defined by (4), and $P_{k+1}$ is a self-scaling matrix. Thus, the parameter $\sigma$ in the condition (12) is a self-scaling factor, which can alter the condition number of the iteration matrix of the conjugate gradient method.

From (30) and (23), we find that if we restrict that

$$\left|y_k^{\mathrm{T}} s_k\right| > \delta \|s_k\|^2, \quad 0 < \delta < 1, \tag{65}$$

then under Lipschitz condition, $|y_k^{\mathrm{T}} s_k| > \delta \|s_k\|^2 \geq \delta / L \|s_k\| \|y_k\|$, i.e., the angle between $y_k$ and $s_k$ is less than $\pi/2$. Thus,

$$d_{k+1}^{\mathrm{T}} g_{k+1} \leq -\lambda_{\min}^{(k+1)} \|g_{k+1}\|^2 \leq -\frac{\sigma\delta}{L^2 + \sigma\delta} \|g_{k+1}\|^2 < 0$$

and

$$\lambda_{\max}^{(k+1)} \leq \omega_k + \sigma \frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k} = \frac{y_k^{\mathrm{T}} y_k s_k^{\mathrm{T}} s_k}{(s_k^{\mathrm{T}} y_k)^2} + \sigma \frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k} \leq \frac{L^2}{\delta^2} + \frac{\sigma}{\delta}.$$

Therefore, according to Lemma 1, if a descent algorithm satisfies the condition (65), then it is globally convergent for nonconvex functions. So, (65) is also an interesting restarting strategy [19]. In fact, (65) is a uniformly convex condition.

Based on (6) and the relationship between the conjugate gradient method and quasi-Newton method, we let

$$\begin{aligned} H_{k+1} &= H_k - \frac{s_k y_k^{\mathrm{T}} H_k + H_k y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \left(\sigma + \frac{y_k^{\mathrm{T}} H_k y_k}{s_k^{\mathrm{T}} y_k}\right) \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} \\ &= H_{k+1}^{\mathrm{BFGS}} + (\sigma - 1) \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}, \end{aligned} \tag{66}$$

where

$$H_{k+1}^{\mathrm{BFGS}} = H_k - \frac{s_k y_k^{\mathrm{T}} H_k + H_k y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \left(1 + \frac{y_k^{\mathrm{T}} H_k y_k}{s_k^{\mathrm{T}} y_k}\right) \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}.$$

So, a new family of quasi-newton method can be obtained from (66), which belongs to Huang's family, but does not belongs to Broyden's family. Hence, it is worth probing further to develop new and more effective unconstrained optimization algorithms. For example, we let $\sigma = c \frac{y_k^{\mathrm{T}} H_k y_k}{s_k^{\mathrm{T}} y_k}$ in (66).

**Acknowledgements** The authors are grateful to the anonymous referees and the Editor-in-chief, Prof. W.W. Hager, for their valuable comments and suggestions on the original version of this paper. The authors also thank N. Andrei for the Fortran code, SCALCG, W.W. Hager and H. Zhang for the C code, CG_DESCENT (Version 5.3) and J.J. Moré for Matlab code, perf.m. Finally, the authors thank PhD Kuiting Zhang (Department of Computer Science, Weifang University) for his help of C language and the Linux system.



## References

1. Al-Baali, M.: Descent property and global convergence of the Fletcher-Reeves method with inexact line-search. IMA J. Numer. Anal. **5**, 121–124 (1985)
2. Andrei, N.: Scaled conjugate gradient algorithms for unconstrained optimization. Comput. Optim. Appl. **38**, 401–416 (2007)
3. Andrei, N.: A scaled BFGS preconditioned conjugate gradient algorithm for unconstrained optimization. Appl. Math. Lett. **20**, 645–650 (2007)
4. Dai, Y.H., Liao, L.Z.: New conjugate conditions and related nonlinear conjugate gradient methods. Appl. Math. Optim. **43**, 87–101 (2001)
5. Dai, Y.H., Yuan, Y.X.: A nonlinear conjugate gradient with a strong global convergence properties. SIAM J. Optim. **10**, 177–182 (1999)
6. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program., Ser. A **91**, 201–213 (2002)
7. Fletcher, R.M., Reeves, C.M.: Function minimization by conjugate gradients. Comput. J. **7**, 149–154 (1964)
8. Gilbert, J.C., Nocedal, J.: Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optim. **2**(1), 21–42 (1992)
9. Gould, N.I.M., Orban, D., Toint, Ph.L.: CUTEr (and SifDec), a constrained and unconstrained testing environment, revisited. ACM Trans. Math. Softw. **29**(4), 373–394 (2003)
10. Hager, W.W., Zhang, H.: A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM J. Optim. **16**, 170–192 (2005)
11. Hager, W.W., Zhang, H.: A survey of nonlinear conjugate gradient methods. Pac. J. Optim. **2**, 35–58 (2006)
12. Hager, W.W., Zhang, H.: Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent. ACM Trans. Math. Softw. **32**(1), 113–137 (2006)
13. Hager, W.W., Zhang, H.: The limited memory conjugate gradient method (2012). www.math.ufl.edu/~hager/papers/CG/lcg.pdf
14. Han, J.Y., Liu, G.H., Yin, H.X.: Convergence of Perry and Shanno's memoryless quasi-Newton method for nonconvex optimization problems. OR Trans. **1**, 22–28 (1997)
15. Hestenes, M.R., Stiefel, E.: Methods of conjugate gradients for solving linear systems. J. Res. Natl. Bur. Stand. **49**(6), 409–439 (1952)
16. Liu, Y., Storey, C.: Efficient generalized conjugate gradient algorithms, part 1: theory. J. Optim. Theory Appl. **69**, 129–137 (1991)
17. Liu, D.Y., Shang, Y.F.: A new Perry conjugate gradient method with the generalized conjugacy condition. In: Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on Issue Date: 10–12 Dec. 2010. doi:10.1109/CISE.2010.5677114
18. Liu, D.Y., Xu, G.Q.: Applying Powell's symmetrical technique to conjugate gradient methods. Comput. Optim. Appl. **49**(2), 319–334 (2011). doi:10.1007/s10589-009-9302-1
19. Liu, D.Y., Xu, G.Q.: A Perry descent conjugate gradient method with restricted spectrum, optimization online, nonlinear optimization (unconstrained optimization), March 2011. http://www.optimization-online.org/DB_HTML/2011/03/2958.html
20. Oren, S.S., Spedicato, E.: Optimal conditioning of self-scaling variable metric algorithms. Math. Program. **10**, 70–90 (1976)
21. Perry, A.: A modified conjugate gradient algorithm. Oper. Res., Tech. Notes **26**(6), 1073–1078 (1978)

22. Polak, E., Ribière, G.: Note sur la convergence de méthodes de directions conjuguées. Rev. Fr. Inform. Rech. Oper. **3**(16), 35–43 (1969)
23. Polyak, B.T.: The conjugate gradient method in extreme problems. USSR Comput. Math. Math. Phys. **9**, 94–112 (1969)
24. Powell, M.J.D.: Restart procedures for the conjugate gradient method. Math. Program. **12**, 241–254 (1977)
25. Shanno, D.F.: Conjugate gradient methods with inexact searches. Math. Oper. Res. **3**, 244–256 (1978)
26. Shanno, D.F.: On the convergence of a new conjugate gradient algorithm. SIAM J. Numer. Anal. **15**, 1247–1257 (1978)